Report on the Second International Workshop on Narrative Extraction from Texts (Text2Story 2019)

Alípio M. Jorge[1,2], Ricardo Campos[1,3], Adam Jatowt[4], Sumit Bhatia[5], Arian Pasquali[1], João Paulo Cordeiro[1,6], Conceição Rocha[1], and Vítor Mangaravite[7]

[1]INESC TEC, Portugal
[2]University of Porto, Portugal
[3]Ci2 - Smart Cities Research Center, Polytechnic Institute of Tomar, Portugal
[4]Kyoto University, Japan
[5]IBM Research AI, India
[6]University of Beira Interior, Portugal
[7]UFMG - Universidade Federal de Minas Gerais, Brazil

Abstract

The Second International Workshop on Narrative Extraction from Texts (Text2Story'19 [http://text2story19.inesctec.pt/]) was held on the 14th of April 2019, in conjunction with the 41st European Conference on Information Retrieval (ECIR 2019) in Cologne, Germany. The workshop provided a platform for researchers in IR, NLP, and design and visualization to come together and share the recent advances in extraction and formal representation of narratives. The workshop consisted of two invited talks, ten research paper presentations, and a poster and demo session. The proceedings of the workshop are available online at http://ceur-ws.org/Vol-2342/

1 Introduction

The continuous growth of social networks such as Facebook and Twitter, together with an ever-increasing presence of traditional news media outlets on the Web, has changed the way information is being generated and consumed. Rather than relying on a few sources of information about an event or a news item topic (e.g., *US and China trade war*), readers now have easy access to the content via multiple, sources (news websites, Facebook posts, Web Archives, etc.) produced by disparate content creators such as journalists, subject matter experts, and social media users. Further, active reader participation also occurs in the comments section of news articles with discussions lasting over days, weeks or possibly, months. Such a stream of continuously evolving information makes it unmanageable and

time-consuming for an interested reader to track, read, and process different sources of information to keep up with all the developments and various aspects of the topic of interest. *Automated narrative extraction from text* offers a compelling approach to this problem and involves identifying the sub-set of interconnected raw documents, extracting the critical narrative/story elements, and representing them in a more adequate final form (e.g., time-lines) that conveys the key points of the *story* in an easy to understand format to the readers or in an intermediate structured formalism that can feed further steps. Although information extraction and natural language processing have made significant progress towards automatic interpretation of texts, the problem of fully identifying and relating the different elements of a narrative present in a document (set) still presents significant unsolved challenges [John *et al.*, 2016].

The workshop brought together a diverse set of participants from academia and industry (about ≈ 40 participants) with expertise in fields such as information retrieval, natural language processing, and design and visualization. The workshop program consisted of two invited keynote talks, ten research papers, and a poster and demos session. The papers presented at the workshop covered diverse aspects of the narrative extraction problem ranging from formal models [Zahid *et al.*, 2019; Staykovski *et al.*, 2019; Metilli *et al.*, 2019; Raiyani *et al.*, 2019; Croce *et al.*, 2019], narrative presentation and visualization [Vani and Antonucci, 2019; Rehm *et al.*, 2019] to different applications such as healthcare [Dirkson *et al.*, 2019], novels [Motwani *et al.*, 2019], and identifying radicalizing narratives [Denaux and Gómez-Pérez, 2019].

The Program Chairs were Alípio Jorge, Ricardo Campos, Adam Jatowt and Sumit Bahtia. The proceedings were setup by our Proceedings Chairs João Paulo Cordeiro and Conceição Rocha. Arian Pasquali was Web Chair and Vitor Mangaravite was in charge of dissemination. We also had Arian Pasquali and Andreas Spitz as Session Chairs.

2   Keynotes

2.1   Iryna Gurevych on Automated Claim Validation

Professor Iryna Gurevych from Technische Universität Darmstadt talked about the crucial task of automatic validation of claims. Claim validations is a difficult task that often requires expert human judgment and Iryna presented an overview of the state-of-the-art on automated claim validation and how these methods compare with manual validation. She also presented an overview of different datasets and benchmarks that can be used for evaluating automated claim validation approaches.

2.2   Miguel Martinez on Automatic Media Analysis

Miguel Martinez, the co-founder and Chief Data Scientist of Signal AI[1], provided a unique industry perspective by describing the real-time online media analysis being carried out at Signal AI. He discussed the challenges in online monitoring and real-time analysis of media sources for various client companies. Companies are interested in the emerging narrative around their products, and media campaigns and large scale monitoring and analysis of news and social media can provide valuable insights to them. The talk touched on various

---

[1] https://www.signal-ai.com/

challenges in developing such solutions and how the IR community can contribute to address those challenges.

3 Research Papers

The papers presented at the workshop covered different facets of narrative extraction from text, which we briefly summarize here.

Narrative Detection and Extraction: Zahid *et al.* [2019] described an approach to automatically segment news story text according to news schema categories such as *action, reaction,* and *history.* Key sentences from each segment can then be sieved together to construct the underlying narrative. Staykovski *et al.* [2019] presented a comparative evaluation of sparse (such as tf-idf) and dense (embeddings based) representations for clustering news articles into topical clusters. Metilli *et al.* [2019] presented their software system for event detection and classification, as part of a larger software framework developed for narrative detection and extraction in digital libraries. Raiyani *et al.* [2019] presented a system for automatic event extraction from Portuguese texts. The proposed approach is based on a pipeline of specialized natural language processing tools; namely, a part-of-speech tagger, a named entity recognizer, a dependency parser, semantic role labeling, and a knowledge extraction module. Croce *et al.* [2019] proposed a machine learning framework combining Hidden Markov Models and SVM to segment drama texts into core dramatic units such as actions, agents and conflicts.

Narrative Presentation and Visualization: Vani and Antonucci [2019] described a framework for presenting visual summaries of narrative texts. Different characters, events, and their interconnections are represented as a graph and provide a succinct summary of the narrative to the end-user. Rehm *et al.* [2019] described their vision of a semantic storytelling system that can identify different characters, events, and plotlines from an input document corpus as a means to present the extracted storylines to the end-users (journalists, researchers, etc.) and aid their respective knowledge processing workflows.

Applications: Motwani *et al.* [2019] presented an interesting application of narrative extraction applied to detective novels where they create evidence summaries for different characters in the novel that can help determine the culprits. Dirkson *et al.* [2019] presented their study of online patient communities that provided insights about the different psycho-linguistic features that can aid in identification of narrative posts in such online discussions. Specifically, they reported that social media posts describing patient narratives are generally written in past tense, and contain health-related words and first-person pronouns. On the other hand, posts containing non-narrative text are often written in future tense, and use emotional support words and second-person pronouns. Denaux and Gómez-Pérez [2019] presented their approach for detecting radicalizing text so that its spread can be minimized. They argued for the need of building such systems by utilizing the research done in social sciences to understand key characteristics of radicalization narratives and presented taxonomies to capture these characteristics. They also described a rule based system that uses these taxonomies to annotate presence of radical narrative text in online documents.

4 Poster and Demo Session

All the accepted research papers were also given an opportunity to showcase their work during the poster session to allow for an informal setting for the workshop participants to discuss the work and enable networking and socializing. In addition, the six following systems were demonstrated during the demo sessions.

1. **TopExNet:** Topic Exploration in Entity-centric Networks of News Streams
   by *Andreas Spitz, Satya Almasian, Michael Gertz* from Heidelberg University
   http://topexnet.ifi.uni-heidelberg.de

2. **TanBih:** News Bias and Media Outlets Profiles
   by *Preslav Nakov* from Qatar Computing Research Institute
   http://www.tanbih.org

3. **ArgumentSearch :** Argumentation Mining Search Engine
   by *Iryna Gurevych* from Technische Universität Darmstadt
   http://argumentsearch.com

4. **Agatha:** Platform for Criminal Police Investigation
   by *Kashyap Raiyani, Teresa Gonçalves, Paulo Quaresma, Vítor Beires Nogueira* from University of Évora
   http://www.agatha-osi.com/

5. **YAKE Keyword Extraction:** A Text Feature Based Automatic Keyword Extraction Method for Single Documents
   by *Ricardo Campos, Vítor Mangaravite, Arian Pasquali, Alípio Jorge, Célia Nunes, Adam Jatowt* from LIAAD INESC TEC / Ci2-IPT / FCUP / UBI / University of Kyoto
   http://yake.inesctec.pt

6. **Tell me stories:** Interactive System for Automatically Generating Temporal Narratives
   by *Arian Pasquali, Vítor Mangaravite, Ricardo Campos, Alípio Jorge, Adam Jatowt* from LIAAD INESC TEC / Ci2-IPT / FCUP/ University of Kyoto
   http://archive.tellmestories.pt, ECIR demo: http://demo.tellmestories.pt

5 Future Directions

This was the second edition of the Text2Story workshop series. Our objective was to bring together the interested participants from different geographies and research expertise to collectively set the agenda for the emerging multi-disciplinary area of *narrative extraction from texts*. In addition to the two workshops [Jorge *et al.*, 2018, 2019a], we have also edited a Special Issue on Narrative Extraction from Texts (Text2Story) in Information Processing & Management (IPM) Journal[2] that provides a good overview of mature research efforts in the field [Jorge *et al.*, 2019b]. We have received very positive response from the research community and the growing participation in the workshops underlines the relevance of the topic. Thus, it is our aim to continue organizing the workshop series in IR and NLP venues to develop the field further.

---

[2]https://www.sciencedirect.com/science/article/abs/pii/S0306457319304455

6 Programme Committee

The following researchers and industry experts have served on the Programme Committee of the Text2Story'19 workshop:

- Álvaro Figueira (University of Porto)
- Arian Pasquali (University of Porto)
- Bruno Martins (IST and INESC-ID - Instituto Superior Técnico, University of Lisbon)
- Conceição Rocha (INESC TEC)
- Daniel Loureiro (University of Porto)
- Daniel Gomes (FCT/Arquivo.pt)
- Dhruv Gupta (Max Planck Institute for Informatics)
- Federico Nanni (University of Mannheim)
- Florian Boudin (Université de Nantes)
- Gaël Dias (Normandie University)
- Gerasimos Lampouras (The University of Sheffield)
- Henrique Lopes Cardoso (University of Porto)
- Jacobo Rouces (University of Gothenburg)
- João Magalhães (New University of Lisbon)
- João Paulo Cordeiro (Universidade da Beira Interior)
- Pablo Gamallo (University of Santiago de Compostela)
- Paulo Quaresma (Universidade de Evora)
- Mark Finlayson (Florida International University)
- Marc Spaniol (Université de Caen Normandie)
- Mengdie Zhuang (The University of Sheffield)
- Miguel Martinez-Alvarez (Signal)
- Nicola Ferro (University of Padova)
- Nina Tahmasebi (University of Gothenburg)
- Nuno Moniz (LIAAD/INESC TEC)
- Udo Kruschwitz (University of Essex)
- Sérgio Nunes (University of Porto)
- Yihong Zhang (Kyoto University)
- Vitor Mangaravite (LIAAD INESC TEC; UFMG)

References

Danilo Croce, Roberto Basili, Vincenzo Lombardo, and Eleonora Ceccaldi. Automatic recognition of narrative drama units: A structured learning approach. In *Proceedings of Text2Story - 2nd Workshop on Narrative Extraction From Texts, co-located with the 41st European Conference on Information Retrieval, Text2Story@ECIR 2019, Cologne, Germany, April 14th, 2019.*, pages 81–88, 2019.

Ronald Denaux and José Manuél Gómez-Pérez. Textual analysis for radicalisation narratives aligned with social sciences perspectives. In *Proceedings of Text2Story - 2nd Workshop on Narrative Extraction From Texts, co-located with the 41st European Conference on Information Retrieval, Text2Story@ECIR 2019, Cologne, Germany, April 14th, 2019.*, pages 39–45, 2019.

Anne Dirkson, Suzan Verberne, and Wessel Kraaij. Narrative detection in online patient communities. In *Proceedings of Text2Story - 2nd Workshop on Narrative Extraction From Texts, co-located with the 41st European Conference on Information Retrieval, Text2Story@ECIR 2019, Cologne, Germany, April 14th, 2019.*, pages 21–28, 2019.

Markus John, Steffen Lohmann, Steffen Koch, Michael Wörner, and Thomas Ertl. Visual analytics for narrative text-visualizing characters and their relationships as extracted from novels. In *VISIGRAPP (2: IVAPP)*, pages 29–40, 2016.

Alípio M. Jorge, Ricardo Campos, Adam Jatowt, and Sérgio Nunes. First international workshop on narrative extraction from text (text2story'18). In *Pasi G., Piwowarski B., Azzopardi L., Hanbury A. (eds). Advances in Information Retrieval. ECIR 2018 (Grenoble, France. March 26 – 29). Lecture Notes in Computer Science, vol 10772*, pages 833–834, 2018.

Alípio M. Jorge, Ricardo Campos, Adam Jatowt, and Sumit Bhatia. Second international workshop on narrative extraction from text (text2story'19). In *Azzopardi L., Stein B., Fuhr N., Mayr P., Hauff C., Hiemstra D. (eds), Advances in Information Retrieval. ECIR'19 (Cologne, Germany. April 14 – 18). Lecture Notes in Computer Science, vol 11438*, pages 389–393, 2019.

Alípio M. Jorge, Ricardo Campos, Adam Jatowt, and Sérgio Nunes. Special issue on narrative extraction from texts (text2story): Preface. In *Information Processing Management an International Journal. Elsevier, Vol 56(5)*, pages 1771–1774, 2019.

Daniele Metilli, Valentina Bartalesi, and Carlo Meghini. Steps towards a system to extract formal narratives from text. In *Proceedings of Text2Story - 2nd Workshop on Narrative Extraction From Texts, co-located with the 41st European Conference on Information Retrieval, Text2Story@ECIR 2019, Cologne, Germany, April 14th, 2019.*, pages 53–61, 2019.

Aditya Motwani, Aayush Naik, and Kamalakar Karlapalem. Extracting evidence summaries from detective novels. In *Proceedings of Text2Story - 2nd Workshop on Narrative Extraction From Texts, co-located with the 41st European Conference on Information Retrieval, Text2Story@ECIR 2019, Cologne, Germany, April 14th, 2019.*, pages 5–11, 2019.

Kashyap Raiyani, Teresa Gonçalves, Paulo Quaresma, and Vítor Beires Nogueira. Automated event extraction model for linked portuguese documents. In *Proceedings of Text2Story - 2nd Workshop on Narrative Extraction From Texts, co-located with the 41st European Conference on Information Retrieval, Text2Story@ECIR 2019, Cologne, Germany, April 14th, 2019.*, pages 13–20, 2019.

Georg Rehm, Karolina Zaczynska, and Julián Moreno. Semantic storytelling: Towards identifying storylines in large amounts of text content. In *Proceedings of Text2Story - 2nd Workshop on Narrative Extraction From Texts, co-located with the 41st European Conference on Information Retrieval, Text2Story@ECIR 2019, Cologne, Germany, April 14th, 2019.*, pages 63–70, 2019.

Todor Staykovski, Alberto Barrón-Cedeño, Giovanni Da San Martino, and Preslav Nakov. Dense vs. sparse representations for news stream clustering. In *Proceedings of Text2Story - 2nd Workshop on Narrative Extraction From Texts, co-located with the 41st European Conference on Information Retrieval, Text2Story@ECIR 2019, Cologne, Germany, April 14th, 2019.*, pages 47–52, 2019.

K. Vani and Alessandro Antonucci. NOVEL2GRAPH: visual summaries of narrative text enhanced by machine learning. In *Proceedings of Text2Story - 2nd Workshop on Narrative Extraction From Texts, co-located with the 41st European Conference on Information Retrieval, Text2Story@ECIR 2019, Cologne, Germany, April 14th, 2019.*, pages 29–37, 2019.

Iqra Zahid, Hao Zhang, Frank Boons, and Riza Batista-Navarro. Towards the automatic analysis of the structure of news stories. In *Proceedings of Text2Story - 2nd Workshop on Narrative Extraction From Texts, co-located with the 41st European Conference on Information Retrieval, Text2Story@ECIR 2019, Cologne, Germany, April 14th, 2019.*, pages 71–79, 2019.